Networks of Probabilistic Events in Discrete Time*

Severino F. Galán and Francisco J. Díez
Dpto. de Inteligencia Artificial
Facultad de Ciencias de la UNED
Paseo senda del rey 9
28040 Madrid (Spain)
{seve, fjdiez}@dia.uned.es

Abstract

The usual methods of applying Bayesian networks to the modeling of temporal processes, such as Dean and Kanazawa's *dynamic Bayesian networks* (DBNs), consist in discretizing time and creating an instance of each random variable for each point in time. We present a new approach called *network of probabilistic events in discrete time* (NPEDT), for temporal reasoning with uncertainty in domains involving probabilistic events. Under this approach, time is discretized and each value of a variable represents the instant at which a certain event may occur. This is the main difference with respect to DBNs, in which the value of a variable V_i represents the state of a real-world property at time t_i. Therefore, our method is more appropriate for temporal fault diagnosis, because only one variable is necessary for representing the occurrence of a fault and, as a consequence, the networks involved are much simpler than those obtained by using DBNs. In contrast, DBNs are more appropriate for monitoring tasks, since they explicitly represent the state of the system at each moment. We also introduce in this paper several types of *temporal noisy gates*, which facilitate the acquisition and representation of uncertain temporal knowledge. They constitute a generalization of traditional canonical models of multicausal interactions, such as the noisy OR-gate, which have been usually applied to static domains. We illustrate the approach with the example domain of modeling the evolution of traffic jams produced on the outskirts of a city, after the occurrence of an event that obliges traffic to stop indefinitely.

Keywords: Bayesian networks, probabilistic temporal reasoning, canonical models, diagnosis, prediction.

*Published in: *International Journal of Approximate Reasoning*, 30:181-202, 2002

1 Introduction

1.1 Bayesian networks

Bayesian networks (BNs) [20, 7] are a probability-based method for representing and reasoning with uncertain knowledge. Each node in a BN is associated with a random variable. In our work all the variables are discrete. Links define probabilistic dependence relations between variables. Formally, a BN is an acyclic directed graph along with a probability distribution for its variables, which satisfies the Markov condition: the probability of any variable V, once determined the values of its parents, is independent of the non-descendants of V. The joint probability over the random variables in the network can be expressed as

$$P(x_1, \ldots, x_n) = \prod_i P(x_i \mid pa(x_i)) \tag{1}$$

where $pa(x_i)$ stands for a configuration of the set of parents of variable X_i. That is to say, the joint probability factorizes in accordance with the network structure, as a consequence of the independence relations codified in it. This result is the basis of the algorithms developed for computing posterior probabilities in BNs.

The first step in designing a BN is the definition of its graph. Then, the conditional probability of each node, given the values of its parents, must be assessed. For every root node, only its a priori probability is needed. Inference consists in both fixing the values of the observed variables and calculating the probability of the unobserved ones.

1.2 BNs and time

BNs have been usually applied without considering an explicit representation of time. However, important efforts have also been made to model temporal processes by means of BNs. These efforts can be classified into three groups: instant-based formalisms, interval-based formalisms, and formalisms based on a representation of time as a continuous variable.

1.2.1 Instant-based formalisms

A way to apply BNs to dynamic domains consists in both discretizing time and creating an instance of each random variable for each point in time. Under the formalism of *dynamic Bayesian networks* (DBNs) [12, 19, 18, 11], initially a static causal model is built. Then, a copy of this model is generated for each instant belonging to a certain temporal range of interest. Finally, links between nodes in adjacent static networks are established. In this way, a DBN obeys the Markov property: the future is conditionally independent of the past given the present. Moreover, the network can perform diagnosis and prediction by means of a standard propagation of evidence. However, if there are long observation sequences, inference becomes impractical in complex DBNs [6].

Among research activities applying DBNs, as defined above, are a model for making judgements concerning persistence of propositions by Dean and Kanazawa [12], a model for sensor validation by Nicholson and Brady [19], a method for reasoning with DBNs by Kjærulff [18], a system for forecasting sleep apnea by Dagum and Galper [10], a qualitative model-based advisory system for therapy planning in gestational diabetes by Hernando *et al.* [16], etc.

1.2.2 Interval-based formalisms

Arroyo-Figueroa and Sucar [3, 4] propose a model called *temporal nodes Bayesian networks* (TNBNs). A TNBN is an extension of a standard BN, in which each temporal node represents an event or a state change of a variable. There is at most one state change for each variable in the temporal range of interest. The value taken on by the variable represents the interval in which the change has occurred. Time is discretized in a finite number of intervals, allowing a different number and duration of intervals for each node (multiple granularity). Each interval for a child node represents possible delays between the occurrence of one of its parent events and the corresponding child state change. Therefore, relative times are used within each temporal node. There is an asymmetry in the way evidence is introduced in the network: the occurrence of an event associated to a node without parents constitutes direct evidence, whereas in the case of a node with parents, several scenarios are possible. When an initial event is detected, its time of occurrence fixes temporally the network. A TNBN permits reasoning about the probability of occurrence of certain events, for diagnosis or prediction, using standard probability propagation techniques developed for BNs. However, this model lacks a formalization of canonical models for temporal processes. Furthermore, each value defined for an effect node, which is associated to a determined time interval, means that the effect has been caused during that interval by only one of its parent events. However, this is not the general case in some domains where evidence about the occurrence of an event can be explained by several of its causes.

Santos and Young [21] develop a technique called *probabilistic temporal network* (PTN) that permits the representation of time constrained causality. BNs provide the probabilistic basis of this model, and management of time is based on Allen's interval system [1] and his thirteen relations. PTNs model processes and the interaction between them. The state of a process is represented by a value at a given time interval. A process can be defined over any number of such intervals. The nodes of the PTN are called *temporal aggregates*. A temporal aggregate consists of a set of states that its associated process can take on, and a set of temporal intervals. Each interval has an associated random variable giving the state of the process over that interval. Different random variables belonging to the same temporal aggregate may be assigned the same value. The edges are the causal temporal relations between aggregates. Edges in the network consist of a disjunctive set of interval relations ($=$, $>$, $<$, etc.) and a schema (OR, XOR, or PASSTHROUGH) to map the random variables of the intervals to a single value. This allows the exact definition of those intervals

	Instants	Intervals	Continuous time
Irreversible processes		TNBN (Arroyo, Sucar)	Network of dates (Berzuini) Continuous time net (Kanazawa)
General processes	DBN (Dean, Kanazawa)	PTN (Santos, Young)	

Table 1: Classification of temporal BNs.

during which the state of one process affects another. The main task is to find the most probable state of the world given some evidence. The most probable explanation is the complete assignment with the greatest joint probability, which is computed by applying the chain rule.

1.2.3 Formalisms based on a representation of time as a continuous variable

Berzuini [5] associates a probability density with each temporal random variable to represent continuous time. When these random variables stand for instantaneous events, they are called *dates*. Attached to each arc, there are conditional distributions specified as conditional intensities, which express how the intensity of the caused event varies as a function of the time elapsed from the occurrence of the causing event. The intensity of an event at time t indicates its propensity to occur just after t, given that it did not occur previously. In the presence of several causes, these causes contribute separate regression functions that combine additively to produce the conditional intensity for the effect. After a suitable discretization of continuous variables involved, exact techniques can be applied for inference. Approximate stochastic techniques may perform without requiring discretization.

Kanazawa [17] presents a formalism called *continuous time net*, which is based on the network of dates proposed by Berzuini. Continuous time nets define a model that uses continuous time and introduces a representation for *fluents* and events. A fluent acts like a dynamic fact that holds over certain time intervals. Conditional influences are represented with simple parametric functions. It is possible to use a wide class of parametric distributions to model event and fact densities.

Table 1 shows a classification of the different BNs for temporal reasoning presented above, according to two parameters: temporal primitive used and permitted number of occurrences for each event.

The rest of this paper is organized as follows. Section 2 deals with canonical models of multicausal interactions. Section 3 details the characteristics of a *network of probabilistic events in discre time*. Section 4 illustrates the application

of the new approach to the management of irreversible processes and compares the result with that obtained by applying DBNs. Section 5 describes an example of application of the approach to a real-world domain. Finally, some concluding remarks are made.

2 Canonical models

In the general case, it is necessary to assign each node in a BN a set of conditional probabilities that grows exponentially with the number of its parents. This complicates the acquisition of the parameters, their storage, and the propagation of evidence. For these reasons, causal interaction models —called canonical models [20]— were developed in order to simplify both BN construction and probability computation. The most famous example is the noisy OR-gate, which requires just one independent parameter per parent.

2.1 Noisy OR-gate

In the noisy OR model [20], each cause X_i (a binary random variable) acts independently of the other causes to produce the effect Y (also a binary random variable). For each X_i, an inhibitory mechanism could prevent this action from taking place, i.e., each present cause may fail to produce the effect with a certain probability. A noisy OR-gate can be decomposed as shown in figure 1. Each auxiliary variable Z_i represents the fact that Y has been produced by X_i. Therefore, $Y = +y$ when $Z_i = +z_i$ for at least one i.

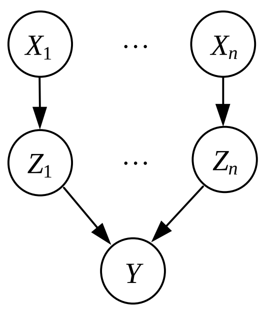

Figure 1: Noisy OR-gate for n causes.

The parameters that define the model are:

$$c_i \equiv P(+z_i \mid +x_i) \tag{2}$$

Put another way, $1 - c_i = P(\neg z_i \mid +x_i)$ is the probability that inhibitor I_i prevents X_i from causing Y. (In a more detailed model, I_i might be represented as a second parent of Z_i.) If X_i is absent, it cannot produce Y; therefore,

$$P(+z_i \mid \neg x_i) = 0 \tag{3}$$

For a certain configuration of X_i's:

$$P(+y \mid \bar{x}) = 1 - \prod_{i \in T_X} (1 - c_i) \tag{4}$$

where T_X is the subset of causes of Y that are present.

In the noisy AND model [20], each parent X_i (a binary random variable) is interpreted as a condition for the effect Y (also a binary random variable). We will only consider the noisy AND without substitutors [14].

2.2 Noisy MAX-gate

The noisy MAX-gate [15, 13, 14] is a generalization for graded variables of the noisy OR-gate. A graded variable E can be either absent or present with g_E degrees of intensity. Usually $E = 0$ means "E is absent" and succeeding integers indicate higher degrees of intensity. This type of causal interaction can be constructed by introducing n auxiliary variables Z_i with the same domain as Y (see figure 1). The parameters of the model are the conditional probabilities:

$$c_y^{x_i} \equiv P(Z_i = y \mid X_i = x_i) \tag{5}$$

The value taken on by Y is the maximum of the z_i's. Therefore, the conditional probability table (CPT) for Y is given by:

$$P(y \mid \bar{x}) = \sum_{\bar{z} \mid \max \bar{z} = y} \prod_i c_{z_i}^{x_i} \tag{6}$$

Figure 2 illustrates equation 6 for a family with two causes, A and B, and one effect, C.

$P(C \mid A{=}a, B{=}b)$	$c^{A=a}_{Z_A=0}$	$c^{A=a}_{Z_A=1}$	$c^{A=a}_{Z_A=2}$	$\dots$	$c^{A=a}_{Z_A=g_C}$
$c^{B=b}_{Z_B=0}$	$C{=}0$	$C{=}1$	$C{=}2$	$\dots$	$C{=}g_C$
$c^{B=b}_{Z_B=1}$	$C{=}1$	$C{=}1$	$C{=}2$	$\dots$	$C{=}g_C$
$c^{B=b}_{Z_B=2}$	$C{=}2$	$C{=}2$	$C{=}2$	$\dots$	$C{=}g_C$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$c^{B=b}_{Z_B=g_C}$	$C{=}g_C$	$C{=}g_C$	$C{=}g_C$	$\dots$	$C{=}g_C$

Figure 2: Noisy MAX-gate for two causes and one effect.

The noisy MIN-gate [13, 14] is a generalization for graded variables of the noisy AND. In the noisy MIN model, the value taken on by Y is the minimum of the z_i's.

2.3 Leaky noisy gates

In real-world applications, it is often unfeasible to enumerate all the possible causes of an effect. In such a case, the non-explicit causes can be implicitly represented in the OR/MAX-gate by a vector of parameters c_y^*, which is the probability that $Y = y$ when the causes explicit in the model are known to be absent. If Y is a binary random variable, it suffices to have one parameter c_{+y}^*.

In conjunctive interaction, the non-explicit conditions can be represented by a leaky vector of parameters h_y^*, giving rise to the leaky noisy AND/MIN-gates.

3 Description of the new approach

In a *network of probabilistic events in discrete time* (NPEDT), each variable represents an event. We consider that each event can happen at most once. Reversible processes are represented by multiple events, one event for each change of state; for instance, the process of turning a light on and off twice consists of four events. Time is discretized by adopting the appropriate temporal unit for each case (seconds, minutes, etc.); therefore, the temporal granularity depends on the particular problem. The value taken on by the variable indicates the time at which the event occurs.

Formally speaking, a temporal random variable V in the network can take on a set of values $v\,[i]$ with $i \in \{a, \ldots, b, never\}$, where a and b are instants defining the limits of the temporal range of interest for V. For example, if V represents "being taken to hospital", $V = v\,[a]$ means that the patient has been taken to hospital at instant a. If the patient is not taken to hospital then $V = v\,[never]$. The links in the network represent temporal causal mechanisms between neighboring nodes. Therefore, each CPT represents the most probable delays between parent events and the corresponding child event. For the case of general dynamic interaction in a family of nodes, giving the CPT involves assessing the probability of occurrence of the child node over time, given any temporal configuration of the parent events. In a family of n parents $X_1, \ldots, X_n$ and one child Y, the CPT is given by

$$P(y\,[t_Y] \mid x_1\,[t_1]\,, \ldots, x_n\,[t_n]) \tag{7}$$

with

$$t_Y \in \{0, \ldots, n_Y, never\}, t_i \in \{0, \ldots, n_i, never\}$$

The joint probability is given by the product of all the CPTs in the network. Any marginal or conditional probability can be derived from the joint probability. For example, if B has happened at t_1 and C at t_2, the a posteriori probability for A is

$$P(a\,[t] \mid b\,[t_1]\,, c\,[t_2]) = \frac{P(a\,[t]\,, b\,[t_1]\,, c\,[t_2])}{P(b\,[t_1]\,, c\,[t_2])} \tag{8}$$

This expression can be used for diagnosis or prediction.

In many domains, the dynamic causal relations have the property of *time invariance*:

$$P(y[t_Y + \triangle t] \mid x_1[t_1 + \triangle t], \ldots, x_n[t_n + \triangle t]) = P(y[t_Y] \mid x_1[t_1], \ldots, x_n[t_n]) \quad (9)$$

If all the CPTs are time-invariant, the network will be time-invariant.

3.1 Temporal nodes without parents

Let A be an event node that may occur at one of instants 0, 1, 2... The probability distribution $P(a[t])$ might be given explicitly. An alternative way of determining this distribution is by defining $P(a^{\dagger}[t])$ as the probability of A being true at t given that it was false at 0, 1, ..., and $t-1$, i.e., the probability that A happens at time t if it has not happened before t. These values can be obtained from a database or estimated by a human expert. As an illustrative example, A could represent the "death caused by an epidemic disease", and $P(a^{\dagger}[t])$ would be the percentage of population dying weekly as a consequence of the disease. (Time could be discretized in weeks.) The probability for temporal node A can be computed by multiplying $P(a^{\dagger}[t])$ by the probability that A has not happened before t:

$$P(a[t]) = P(a^{\dagger}[t]) \cdot \left(1 - \sum_{t'=0}^{t-1} P(a[t'])\right) \quad \forall t > 0 \quad (10)$$

In the area of survival analysis (cf. [9, 22]), $P(a^{\dagger}[t])$ is called *hazard function* and the second factor in equation 10 is the *survivor function* for event A. Equation 10 is useful when $P(a^{\dagger}[t])$ is a constant and does not depend on t. If $P(a^{\dagger}[t])$ $= k$, equation 10 leads to:

$$P(a[t]) = (1-k) \cdot P(a[t-1]) \quad \forall t > 0 \quad (11)$$

Proof 1

$$\begin{aligned} P(a[t]) &= k \cdot \left(1 - \left[\sum_{t'=0}^{t-2} P(a[t'])\right] - P(a[t-1])\right) \\ &= P(a[t-1]) - k \cdot P(a[t-1]) \\ &= (1-k) \cdot P(a[t-1]) \end{aligned}$$

From equation 11,

$$P(a[t]) = (1-k)^t \cdot k \quad (12)$$

Therefore, $P(a[t])$ is a probability distribution with exponential decay. Since

$$(1-k)^t = 1 - kt - \frac{1}{2}k^2t^2 + \ldots \quad (13)$$

when $kt \ll 1$ for the temporal range of interest, $(1-k)^t \approx 1$; therefore, $P(a[t]) \approx k$. We then have time invariance for node A.

3.2 Node with one parent

Let us consider the network in figure 3. The temporal ranges of interest for events A and B are $\{0,\ldots,t_A\}$ and $\{0,\ldots,t_B\}$, respectively.

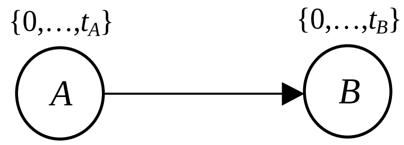

Figure 3: Temporal network with one parent and one child.

$P(b\,[j] \mid a\,[i])$ is the probability that B happens at j when A has happened at i. The CPT for link $A \to B$ can be as general as possible, permitting any delay between A and B, with probabilities varying over time for each particular delay (see table 2). When $j < i$, $P(b\,[j] \mid a\,[i]) = 0$ because the effect cannot precede the cause. When $i = never$, $P(b\,[j] \mid a\,[i]) = 0$ as well.

$B \setminus A$	$a\,[0]$	$a\,[1]$	$a\,[2]$	$a\,[never]$
$b\,[0]$	0.5	0	0	0
$b\,[1]$	0.1	0.3	0	0
$b\,[2]$	0.1	0.05	0.2	0
$b\,[3]$	0.1	0.02	0.2	0
$b\,[never]$	0.2	0.63	0.6	1

Table 2: A general CPT for t_A=2 and t_B=3.

If we had a time-invariant causal relation for arc $A \to B$ then

$$P(b\,[j+\triangle t] \mid a\,[i+\triangle t]) = P(b\,[j] \mid a\,[i]) \tag{14}$$

with

$$j, j+\triangle t \in \{0,\ldots,t_B\} \quad and \quad i, i+\triangle t \in \{0,\ldots,t_A\}$$

Under this assumption, we only need to specify a probability for each delay. Two special cases of time invariance are worth being taken into account:

- **Delays limited in time**

  In this case, there is a finite number of possible delays between the occurrences of the parent event and the child event. For example, once we know that A has taken place, the probability of B at that instant could be 0.5, and 0.1 one instant later:

$$P(b\,[j] \mid a\,[i]) = \begin{cases} 0.5 & if \quad j = i \\ 0.1 & if \quad j = i+1 \\ 0 & otherwise \end{cases} \tag{15}$$

$B \setminus A$	$a[0]$	$a[1]$	$a[2]$	$a[never]$
$b[0]$	0.5	0	0	0
$b[1]$	0.1	0.5	0	0
$b[2]$	0	0.1	0.5	0
$b[3]$	0	0	0.1	0
$b[never]$	0.4	0.4	0.4	1

Table 3: Time-invariant CPT with two delays for t_A=2 and t_B=3.

The conditional probabilities for arc $A \rightarrow B$ appear in table 3. This table shows that the possible delays between cause and effect are 0 and 1, with associated probabilities 0.5 and 0.1, respectively. These probabilities can be estimated by a human expert or obtained from a database by taking into account the delay between A and B.

- **Exponential decay**

  In the same way that the prior probabilities of a root node A can be given directly or through the parameters $P(a^{\dagger}[t])$, we can also give $P(b[t] \mid a[i])$ directly, as we have done so far, or by defining the parameters $P(b^{\dagger}[t] \mid a[i])$. $P(b^{\dagger}[t] \mid a[i])$ represents the probability that B happens at instant t given that it has not happened before t, and that A has taken place at i. Since A is the only cause of B,

$$P(b[t] \mid a[i]) = \begin{cases} P(b^{\dagger}[t] \mid a[i]) \cdot \left(1 - \sum_{t'=i}^{t-1} P(b[t'] \mid a[i])\right) & t > i \\ P(b^{\dagger}[t] \mid a[i]) & t = i \\ 0 & t < i \end{cases} \quad (16)$$

  Equation 16 is mathematically equivalent to equation 10. If $\forall t \geq i$ $P(b^{\dagger}[t] \mid a[i]) = k$ then we have exponential decay for the probability distribution of B, given A at instant i:

$$P(b[t] \mid a[i]) = \begin{cases} k \cdot (1-k)^{t-i} & t \geq i \\ 0 & t < i \end{cases} \quad (17)$$

3.3 Canonical models and time

If we consider a family of nodes with n parents and divide our temporal range of interest into i instants, in the general case $i \cdot (i+1)^n$ independent conditional probabilities have to be assessed to complete the CPT associated to the child node. If we assume time invariance, this number changes into $i \cdot (i+1)^n - (i-1)^{n+1}$ independent parameters. In real-world applications, it is difficult to find a human expert or a database that allows us to create such tables, due to the exponential growth of the set of required parameters with the number of parents. For this reason, a formalization for temporal domains of traditional canonical models turns out to be convenient. In these models, the set of required independent conditional probabilities for a family of nodes is linear with the number of parents.

3.3.1 Temporal noisy OR-gate

We are dealing with domains that can be modeled by associating random variables to events. In the static case, the noisy OR-gate appropriately reproduces the kind of interactions in which both the presence of one cause is sufficient to produce the effect and this causal mechanism is independent of the rest of the causes. For temporal processes, additional questions should be taken into account, as shown below.

Let us consider a network with n causes $X_1, \ldots, X_n$ and one effect Y. The temporal ranges for these nodes are, respectively, $\{0, \ldots, t_{X_1}\}, ..., \{0, \ldots, t_{X_n}\}$, and $\{0, \ldots, t_Y\}$. Each parameter c_i that appeared in the static case (see section 2.1), now separates into parameters

$$c_{y[k]}^{x_i[j_i]} \equiv P(Z_i = y\,[k] \mid X_i = x_i\,[j_i]) \tag{18}$$

with

$$j_i \in \{0, \ldots, t_{X_i}, never\}, k \in \{0, \ldots, t_Y, never\}$$

allowing different delays between cause and effect. The type of relation between X_i and Z_i was described in section 3.2. We are interested in calculating the probability of Y at any instant given evidence about its causes, as indicated in figure 4, where $n = 2$, $X_1 = x_1\,[i_1]$, and $X_2 = x_2\,[i_2]$. The general reasoning followed in figure 4 establishes that if X_1 causes Y to be true at t_1, and X_2 causes Y to be true at t_2, then Y becomes true at $\min(t_1, t_2)$, since every event can happen only once. Therefore, the *temporal noisy OR-gate* represents the case in which the effect is present as soon as any of its causes provokes it to be present and can be represented by a non-temporal noisy MIN.

$P(y[t] \mid x_1[i_1], x_2[i_2])$	$c_{y[never]}^{x_1[i_1]}$	$c_{y[n_Y]}^{x_1[i_1]}$	$\ldots$	$c_{y[1]}^{x_1[i_1]}$	$c_{y[0]}^{x_1[i_1]}$
$c_{y[never]}^{x_2[i_2]}$	$y[never]$	$y[n_Y]$	$\ldots$	$y[1]$	$y[0]$
$c_{y[n_Y]}^{x_2[i_2]}$	$y[n_Y]$	$y[n_Y]$	$\ldots$	$y[1]$	$y[0]$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$c_{y[1]}^{x_2[i_2]}$	$y[1]$	$y[1]$	$\ldots$	$y[1]$	$y[0]$
$c_{y[0]}^{x_2[i_2]}$	$y[0]$	$y[0]$	$\ldots$	$y[0]$	$y[0]$

Figure 4: Temporal noisy OR-gate for two causes.

From figure 4, a noisy MAX-gate leads to a temporal noisy OR-gate by associating increasing intensity degrees to decreasing temporal indices ($never < n_Y < \ldots < 1 < 0$). Note that value "*never*" is analogous to value "absent" of a graded variable. Therefore, a temporal noisy OR-gate can be modeled through a noisy MAX-gate by ordering temporal values from future to past.

It is interesting to study the case where each relation between X_i and Z_i is characterized by an exponential decay probability for Z_i, given X_i at a certain instant:

$$P(z_i\,[t] \mid x_i\,[t_i]) = \begin{cases} k_i \cdot (1-k_i)^{t-t_i} & t \geq t_i \\ 0 & t < t_i \end{cases} \tag{19}$$

Figure 5, where we suppose that $i_1 < i_2$, is the result of transforming figure 4 for this particular case.

$P(y[t] \mid x_1[i_1], x_2[i_2])$	$\dots$	$k_1(1-k_1)^{i_2-i_1+1}$	$k_1(1-k_1)^{i_2-i_1}$	$\dots$	$k_1(1-k_1)$	k_1	$\dots$	0	0
$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$k_2(1-k_2)$	$\dots$	$y[i_2+1]$	$y[i_2]$	$\dots$	$y[i_1+1]$	$y[i_1]$	$\dots$	$y[1]$	$y[0]$
k_2	$\dots$	$y[i_2]$	$y[i_2]$	$\dots$	$y[i_1+1]$	$y[i_1]$	$\dots$	$y[1]$	$y[0]$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0	$\dots$	$y[i_1+1]$	$y[i_1+1]$	$\dots$	$y[i_1+1]$	$y[i_1]$	$\dots$	$y[1]$	$y[0]$
0	$\dots$	$y[i_1]$	$y[i_1]$	$\dots$	$y[i_1]$	$y[i_1]$	$\dots$	$y[1]$	$y[0]$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
0	$\dots$	$y[1]$	$y[1]$	$\dots$	$y[1]$	$y[1]$	$\dots$	$y[1]$	$y[0]$
0	$\dots$	$y[0]$	$y[0]$	$\dots$	$y[0]$	$y[0]$	$\dots$	$y[0]$	$y[0]$

Figure 5: Temporal noisy OR-gate for two causes and conditional probabilities with exponential decay.

From figure 5,

$$P(y\,[t] \mid x_1\,[i_1]\,, x_2\,[i_2]) = \begin{cases} 0 & 0 \leq t < i_1 \\ k_1 \cdot (1-k_1)^{t-i_1} & i_1 \leq t < i_2 \\ k_{OR} \cdot (1-k_1)^{t-i_1}(1-k_2)^{t-i_2} & i_2 \leq t < \infty \end{cases} \tag{20}$$

with

$$k_{OR} = 1 - (1-k_1) \cdot (1-k_2)$$

Proof 2 *For $0 \leq t < i_1$, Y cannot be produced either by X_1 or by X_2. Therefore, $P(y\,[t] \mid x_1\,[i_1]\,, x_2\,[i_2]) = 0$.*

For $i_1 \leq t < i_2$, Y can only be caused by X_1. This corresponds to the case presented in section 3.2. Anyhow, from figure 5 we have

$$\begin{aligned} P(y\,[t] \mid x_1\,[i_1]\,, x_2\,[i_2]) &= k_1(1-k_1)^{t-i_1} \sum\nolimits_{t'=0}^{\infty} k_2(1-k_2)^{t'} \\ &= k_1(1-k_1)^{t-i_1} k_2 \frac{1}{1-(1-k_2)} \\ &= k_1(1-k_1)^{t-i_1} \end{aligned}$$

For $i_2 \le t < \infty$, there are two possible causes for Y: X_1 and X_2. From figure 5,

$$\begin{aligned}
P(y[t] \mid x_1[i_1], x_2[i_2]) &= k_1(1-k_1)^{t-i_1} k_2(1-k_2)^{t-i_2} + \\
&+ k_1(1-k_1)^{t-i_1} \sum\nolimits_{t'=t-i_2+1}^{\infty} k_2(1-k_2)^{t'} + \\
&+ k_2(1-k_2)^{t-i_2} \sum\nolimits_{t'=t-i_1+1}^{\infty} k_1(1-k_1)^{t'} \\
&= k_1(1-k_1)^{t-i_1} k_2(1-k_2)^{t-i_2} + \\
&+ k_1(1-k_1)^{t-i_1} \left(1 - \sum\nolimits_{t'=0}^{t-i_2} k_2(1-k_2)^{t'}\right) + \\
&+ \ k_2(1-k_2)^{t-i_2} \left(1 - \sum\nolimits_{t'=0}^{t-i_1} k_1(1-k_1)^{t'}\right) \\
&= k_1(1-k_1)^{t-i_1} k_2(1-k_2)^{t-i_2} + \\
&+ k_1(1-k_1)^{t-i_1} (1-k_2)^{t-i_2+1} + \\
&+ k_2(1-k_2)^{t-i_2} (1-k_1)^{t-i_1+1} \\
&= k_{OR}(1-k_1)^{t-i_1}(1-k_2)^{t-i_2}
\end{aligned}$$

Note that for $i_1 > i_2$, equation 20 remains valid provided that we interchange in it subscripts "1" and "2". When $i_1 = i_2 = i$, $P(y[t] \mid x_1[i], x_2[i]) = k_{OR} \cdot \{1 - [1-(1-k_1)(1-k_2)]\}^{t-i}$, which is like having only one cause with a decay constant $k_{OR} = 1-(1-k_1)(1-k_2)$.

3.3.2 Temporal noisy AND-gate

The *temporal noisy AND-gate* represents the case in which the effect is present as soon as all its conditions have permitted it to be present. Under this type of interaction, if X_1 permits Y to be true at t_1, and X_2 at t_2, we consider that Y becomes true at $\max(t_1, t_2)$. Therefore, the temporal noisy AND-gate can be represented by a non-temporal noisy MAX.

A noisy MIN-gate leads to a temporal noisy AND-gate by associating increasing intensity degrees to decreasing temporal indices. Consequently, a temporal noisy AND-gate can be modeled though a noisy MIN-gate by ordering the temporal values from future to past.

In canonical model applications, disjunctive interactions (OR gate) appear much more often than conjunctive ones. That is because we are mainly interested in modeling the evolution of failures, anomalies, or malfunctions in a system, either in the past (diagnosis) or in the future (prediction). In this kind of domains, disjunctive interaction is directly related to our intuitive notion of causality. (For an anomaly to appear, only one of its causes is needed.) Conversely, in other domains we are interested in the evolution of a system from a state of malfunction to another of normality. In this case, event nodes represent processes of recovery that interact conjunctively. For example, after a car accident resulting in multiple injuries to a person, we could model the process of

recovery as shown in figure 6. Each variable X_i represents the event “the patient starts to be treated for injury i”. Variable Y represents “complete recovery”. All these variables interact through a temporal noisy AND-gate because each X_i is a condition for Y. Of course, if the X_i’s are not independent, the model must contain links among them or common ancestors of these nodes. The temporal range of interest is $\{0, \ldots, m\}$, where $t = 0$ is the time the accident occurs and $t = m$ is an arbitrary time point. We introduce for each condition X_i, one auxiliary variable Z_i representing “recovery from injury i”.

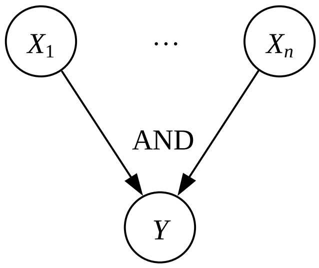

Figure 6: Network for modeling the process of complete recovery from an accident.

The parameters needed to complete the model are the conditional probabilities:

$$h_{y[k]}^{x_i[j_i]} \equiv P(Z_i = y\,[k] \mid X_i = x_i\,[j_i]) \tag{21}$$

with

$$j_i, k \in \{0, \ldots, m\} \quad and \quad i \in \{1, \ldots, n\}$$

and the prior probabilities:

$$P(x_i\,[j_i]) \quad \forall j_i \in \{0, \ldots, m\}, \forall i \in \{1, \ldots, n\} \tag{22}$$

The conditional probabilities give us an idea of the most probable durations of successful treatments, and the prior probabilities indicate the times treatments usually start to be applied after the accident. (Some treatments can be applied just after the accident occurrence, others can only be applied in a hospital, etc.) The event “complete recovery” will be true as soon as recovery from the last injury has taken place.

3.3.3 Temporal leaky noisy gates

Under the hypotheses introduced by Díez and Druzdzel for leaky models [14], the non-explicit causes of a node Y can be grouped together and represented by a vector of parameters. If the temporal range for Y is $\{0, \ldots, t_Y\}$, we only need to give the parameters

$$c_{y[i]}^{*} \quad \forall i \in \{0, \ldots, t_Y\} \tag{23}$$

Therefore, a *temporal leaky noisy OR-gate* can be modeled by means of a leaky noisy MAX-gate. To that end, temporal indices must be ordered from future to past.

The non-explicit conditions in a conjunctive model can be represented through a vector of parameters

$$h^*_{y[i]} \quad \forall i \in \{0, \ldots, t_Y\} \tag{24}$$

A *temporal leaky noisy AND-gate* can be modeled through a leaky noisy MIN-gate by ordering temporal indices from future to past.

4 Comparison with DBNs applied to irreversible processes

Consider the network shown in figure 7 where prior probabilities for root nodes and conditional probabilities are given directly.

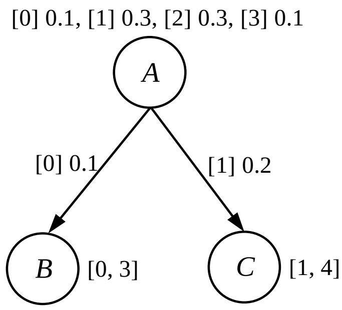

Figure 7: Network with one parent and two children.

In the NPEDT approach, the graph of the network does not change. Other parts of the network are:

- Nodes: A, B, and C.
- Values for the nodes: $A = \{a\,[0],\ a\,[1],\ a\,[2],\ a\,[3],\ a\,[never]\}$, $B = \{b\,[0],\ b\,[1],\ b\,[2],\ b\,[3],\ b\,[never]\}$, and $C = \{c\,[1],\ c\,[2],\ c\,[3],\ c\,[4],\ c\,[never]\}$
- A priori probabilities for root nodes: $P(a\,[0]) = 0.1$, $P(a\,[1]) = 0.3$, $P(a\,[2]) = 0.3$, $P(a\,[3]) = 0.1$, and $P(never) = 0.2$.
- The CPTs, constructed as delays, appear in tables 4 and 5.

In DBNs, the way to deal with temporal information is different in comparison with our approach. Kanazawa's proposal [17] for representing irreversible events by means of DBNs consists in associating a binary node to each possible occurrence of the event at a point in time. *Memory nodes* (cf. [17], section

$B \setminus A$	$a\,[0]$	$a\,[1]$	$a\,[2]$	$a\,[3]$	$a\,[never]$
$b\,[0]$	0.1	0	0	0	0
$b\,[1]$	0	0.1	0	0	0
$b\,[2]$	0	0	0.1	0	0
$b\,[3]$	0	0	0	0.1	0
$b\,[never]$	0.9	0.9	0.9	0.9	1

Table 4: CPT for arc $A \rightarrow B$.

$C \setminus A$	$a\,[0]$	$a\,[1]$	$a\,[2]$	$a\,[3]$	$a\,[never]$
$c\,[1]$	0.2	0	0	0	0
$c\,[2]$	0	0.2	0	0	0
$c\,[3]$	0	0	0.2	0	0
$c\,[4]$	0	0	0	0.2	0
$c\,[never]$	0.8	0.8	0.8	0.8	1

Table 5: CPT for arc $A \rightarrow C$.

4.3.1) prevent each event from taking place at two different points in time. A memory node is true when its associated event is true or when that event has already happened in the past (see figure 8). Figure 8 shows the transformation of the network in figure 7 into a DBN. This DBN is formed by:

- Nodes: A_0, A_1, A_2, A_3, B_0, B_1, B_2, B_3, C_1, C_2, C_3, and C_4. In addition, for every event N_i there is one node $\mathrm{M}(N_i)$ that is true provided that N has taken place at instant i or earlier.
- Values for each node: $\{$*true*, *false*$\}$
- A priori probabilities for root nodes: $P(A_0 = true) = 0.1$, and $P(A_0 = false) = 0.9$.
- CPTs: In table 6 we show the CPTs for the network in figure 8.

Although both types of networks lead to identical posterior probabilities given some evidence, the network in figure 8 is much more complex. In general, the formalism of DBNs leads to networks with a high structural complexity. In the case of irreversible processes, this complexity disappears by adopting the method presented in this paper. Therefore, if time is discretized in a set of temporal points, domains involving irreversible processes are better modeled through an NPEDT. However, when the same process may happen multiple times, DBNs are a better option.

Finally, an advantage of NPEDTs over DBNs is that the former are not restricted to Markovian processes. Note that in a DBN, links connect either nodes within the same static network or between adjacent static networks.

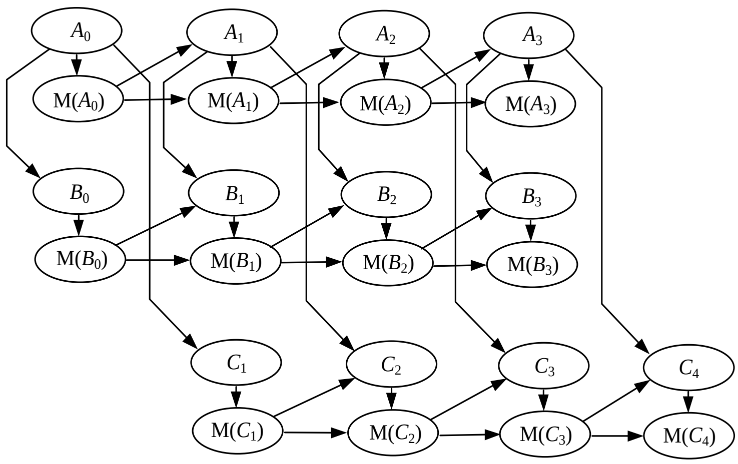

Figure 8: DBN with memory nodes for the network in figure 7.

$A_i\backslash \mathrm{M}(A_{i-1})$	*true*	*false*
true	0	$P(a^\dagger\,[i])$
false	1	$1 - P(a^\dagger\,[i])$

$A_2\backslash \mathrm{M}(A_1)$	*true*	*false*
true	0	$\frac{0.3}{1-0.1-0.3}$
false	1	$1/2$

$\backslash \mathrm{M}(X_{i-1})$	*true*		*false*	
$\mathrm{M}(X_i)\backslash X_i$	*true*	*false*	*true*	*false*
true	1	1	1	0
false	0	0	0	1

$\backslash \mathrm{M}(B_{i-1})$	*true*		*false*	
$B_i\backslash A_i$	*true*	*false*	*true*	*false*
true	0	0	0.1	0
false	1	1	0.9	1

$\backslash \mathrm{M}(C_{i-1})$	*true*		*false*	
$C_i\backslash A_{i-1}$	*true*	*false*	*true*	*false*
true	0	0	0.2	0
false	1	1	0.8	1

Table 6: CPTs of the DBN associated to the example.

5 An example

We aim to model the evolution of traffic jams produced on the outskirts of a city, after the occurrence of an event (road accident, natural catastrophe, etc.) that obliges traffic to stop *indefinitely*. We are interested in the prior period to the start of the working day, since there is a higher density of vehicles. In figure 9, nodes C_1 and C_2 represent two different destination points for commuters, which are located in the city center. Each node P_1, P_2, or P_3 could be a crossroads, a commuter town..., i.e., they are not destination points.

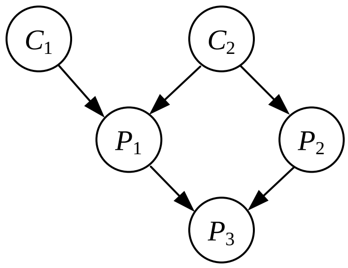

Figure 9: Outskirts of the city.

Next we illustrate the application of the NPEDT approach to the domain described above. Each node represents the event "traffic jam appearance" at that place. Arrows indicate the direction of traffic jam propagation. Traffic jams propagate in the opposite direction to traffic. Our temporal range of interest is from 8:10 a.m. to 9:00 a.m., which we divide into ten-minute intervals. Each family of nodes in the network interacts through a temporal noisy OR model; for example, a traffic jam may appear at P_3 if it has appeared before at either P_1 or P_2. $P_3 = p_3\,[8{:}30]$ means that a traffic jam has initiated at P_3 between 8:21 and 8:30 in the morning. We suppose that the initial event causing the problems, prevents the traffic from being restored within our temporal range of interest (from 8:10 a.m. to 9:00 a.m.).

Table 7 shows the prior probabilities for C_1 and C_2. For instance, $P(c_1\,[8{:}40])$ is the probability of occurrence of an event causing an indefinite traffic jam at C_1 between 8:31 and 8:40 in the morning. The probabilities presented throughout this section are arbitrary ones. Anyway, the city itself is the best database

$c_1\,[8{:}20]$	$c_1\,[8{:}30]$	$c_1\,[8{:}40]$	$c_1\,[8{:}50]$	$c_1\,[9{:}00]$	$c_1\,[never]$
0.0001	0.0005	0.001	0.0025	0.003	0.9929

$c_2\,[8{:}20]$	$c_2\,[8{:}30]$	$c_2\,[8{:}40]$	$c_2\,[8{:}50]$	$c_2\,[9{:}00]$	$c_2\,[never]$
0.0002	0.0005	0.0015	0.003	0.0035	0.9913

Table 7: Prior probabilities for C_1 and C_2.

from which to obtain these parameters. Note that the probability of traffic jam increases as time evolves. This is because the working day usually begins at 9:00 a.m. and the probability of an accident depends on the density of vehicles. The CPTs can be constructed from the parameters in tables 8, 9, and 10.

The parameters in the network show that as we approach either the city center or 9:00 a.m., the probability of traffic jam propagation between neighboring nodes increases, as well as the probability that this propagation takes place in a shorter period of time. Note that the distances between P_1 and C_2, and P_2 and C_2, are much longer than for the rest of the arcs. There are non-explicit causes in the model. (For example, an accident could cause a traffic jam in a peripheral node, while there is a normal situation in the city center.)

Table 11 shows the posterior probabilities for $P_2 = p_2\,[9{:}00]$. From this table, $c_2\,[8{:}20]$ and $c_2\,[8{:}30]$ are the most probable explanations for the observed evidence. If we also knew that $P_1 = p_1\,[8{:}30]$ then C_2 would not explain the evidence on P_1 and, as a result, there would be an increase in the probability of $c_1\,[8{:}20]$ and $c_1\,[8{:}30]$ (see table 12).

6 Conclusions

The process of computing posterior probabilities in BNs is NP-hard [8]. This complexity becomes particularly problematic in large models such as those that arise when modeling temporal processes by *dynamic Bayesian networks* (DBNs). We have presented a new method called *network of probabilistic events in discrete time* (NPEDT), for handling temporal information through BNs. Our model is similar to Arroyo-Figueroa and Sucar's *temporal nodes Bayesian networks* (TNBNs) [3, 4, 2] in that variables represent events and each value of a variable represents the time at which an event takes place. A minor difference is that their model represents time by means of intervals of different duration while our model assumes discrete time. A limitation of Arroyo-Figueroa and Sucar's model is that, when a variable represents an effect of several possible causes, each interval of this variable is associated with only one of the causes. We overcome this limitation by introducing *temporal noisy gates*, which are a generalization of some canonical models of causal interaction.

When compared to DBNs, our method is more appropriate for fault diagnosis in temporal domains, because it uses only one variable to represent the occurrence of a fault and the networks involved are therefore much simpler than those obtained by applying DBNs. In contrast, DBNs are more appropriate for monitoring tasks, since they explicitly represent the state of the system at each moment.

Acknowledgments

This research was supported by the Spanish CICYT, under grants TIC97-0604 and TIC-97-1135-C04-04.

P_1	c_1 [8:20]	c_1 [8:30]	c_1 [8:40]	c_1 [8:50]	c_1 [9:00]	leaky
p_1 [8:20]	0.15	0	0	0	0	0.00005
p_1 [8:30]	0.15	0.15	0	0	0	0.0002
p_1 [8:40]	0.15	0.15	0.2	0	0	0.0005
p_1 [8:50]	0.2	0.2	0.2	0.25	0	0.0011
p_1 [9:00]	0.25	0.3	0.3	0.35	0.5	0.0014
p_1 [*never*]	$1-0.9$	$1-0.8$	$1-0.7$	$1-0.6$	$1-0.5$	0.99675
	c_2 [8:20]	c_2 [8:30]	c_2 [8:40]	c_2 [8:50]	c_2 [9:00]	
p_1 [8:20]	0	0	0	0	0	
p_1 [8:30]	0	0	0	0	0	
p_1 [8:40]	0	0	0	0	0	
p_1 [8:50]	0.1	0	0	0	0	
p_1 [9:00]	0.13	0.14	0	0	0	
p_1 [*never*]	$1-0.23$	$1-0.14$	1	1	1	

Table 8: Parameters for P_1.

$P_2 \backslash C_2$	c_2 [8:20]	c_2 [8:30]	c_2 [8:40]	c_2 [8:50]	c_2 [9:00]	leaky
p_2 [8:20]	0	0	0	0	0	0.00004
p_2 [8:30]	0	0	0	0	0	0.0003
p_2 [8:40]	0	0	0	0	0	0.0004
p_2 [8:50]	0.13	0	0	0	0	0.0012
p_2 [9:00]	0.15	0.15	0	0	0	0.0015
p_2 [*never*]	$1-0.28$	$1-0.15$	1	1	1	0.99656

Table 9: Parameters for P_2.

P_3	p_1 [8:20]	p_1 [8:30]	p_1 [8:40]	p_1 [8:50]	p_1 [9:00]	leaky
p_3 [8:20]	0.12	0	0	0	0	0.00002
p_3 [8:30]	0.12	0.14	0	0	0	0.0001
p_3 [8:40]	0.14	0.15	0.15	0	0	0.0003
p_3 [8:50]	0.18	0.19	0.2	0.2	0	0.0006
p_3 [9:00]	0.2	0.22	0.25	0.3	0.4	0.0008
p_3 [*never*]	$1-0.76$	$1-0.7$	$1-0.6$	$1-0.5$	$1-0.4$	0.99818
	p_2 [8:20]	p_2 [8:30]	p_2 [8:40]	p_2 [8:50]	p_2 [9:00]	
p_3 [8:20]	0.1	0	0	0	0	
p_3 [8:30]	0.12	0.12	0	0	0	
p_3 [8:40]	0.15	0.15	0.15	0	0	
p_3 [8:50]	0.18	0.2	0.2	0.19	0	
p_3 [9:00]	0.2	0.24	0.26	0.29	0.39	
p_3 [*never*]	$1-0.75$	$1-0.71$	$1-0.61$	$1-0.48$	$1-0.39$	

Table 10: Parameters for P_3.

	8:20	8:30	8:40	8:50	9:00	*never*
C_1	0.0001	0.0005	0.001	0.0025	0.003	0.9929
C_2	**0.01879**	**0.04704**	0.0014	0.0028	0.00327	0.92667
P_1	0.00006	0.00028	0.00078	0.00391	0.01319	0.98174
P_3	0.00002	0.00014	0.00047	0.0016	0.3937	0.60404

Table 11: Posterior probabilities for $P_2 = p_2$ [9:00].

	8:20	8:30	8:40	8:50	9:00	*never*
C_1	**0.05177**	**0.25892**	0.00068	0.00172	0.00206	0.68481
C_2	0.01879	0.04704	0.0014	0.0028	0.00327	0.92667
P_3	0.00002	0.14008	0.15019	0.19023	0.3368	0.18266

Table 12: Posterior probabilities for $P_1 = p_1$ [8:30] and $P_2 = p_2$ [9:00].

References

[1] J. F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.

[2] G. Arroyo-Figueroa. *Razonamiento Probabilístico con Nodos Temporales y su Aplicación al Diagnóstico y Predicción de Eventos*. PhD thesis, Instituto Tecnológico y de Estudios Superiores de Monterrey, Cuernavaca, Mexico, 1999. In Spanish.

[3] G. Arroyo-Figueroa and L. E. Sucar. A temporal Bayesian network for diagnosis and prediction. In *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI'99)*, pages 13–20, Stockholm, Sweden, 1999. Morgan Kaufmann, San Francisco, CA.

[4] G. Arroyo-Figueroa, L. E. Sucar, and A. Villavicencio. Probabilistic temporal reasoning and its application to fossil power plant operation. *Expert Systems with Applications*, 15:317–324, 1998.

[5] C. Berzuini. Representing time in causal probabilistic networks. In M. Henrion, R. D. Shachter, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5*, pages 15–28. Elsevier Science Publishers, Amsterdam, 1990.

[6] J. Binder, K. Murphy, and S. Russell. Space-efficient inference in dynamic probabilistic networks. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI–97)*, pages 1292–1296, Nagoya, Japan, 1997.

[7] E. Castillo, J. M. Gutiérrez, and A. S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer Verlag, New York, 1997.

[8] G. F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393–405, 1990.

[9] D. R. Cox and D. Oakes. *Analysis of Survival Data.* Chapman and Hall, 1984.

[10] P. Dagum and A. Galper. Forecasting sleep apnea with dynamic network models. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence (UAI'93)*, pages 64–71, Washington D.C., 1993. Morgan Kaufmann, San Francisco, CA.

[11] P. Dagum, A. Galper, and E. Horvitz. Dynamic network models for forecasting. In *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence (UAI'92)*, pages 41–48, Stanford University, 1992. Morgan Kaufmann, San Francisco, CA.

[12] T. Dean and K. Kanazawa. A model for reasoning about persistence and causation. *Computational Intelligence*, 5:142–150, 1989.

[13] F. J. Díez. Parameter adjustment in Bayes networks. The generalized noisy OR–gate. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence (UAI'93)*, pages 99–105, Washington D.C., 1993. Morgan Kaufmann, San Francisco, CA.

[14] F. J. Díez and M. Druzdzel. Canonical probabilistic models for knowledge engineering. Technical Report, Decision Systems Laboratory, University of Pittsburgh, 2002. In preparation.

[15] M. Henrion. Some practical issues in constructing belief networks. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*, pages 161–173. Elsevier Science Publishers, Amsterdam, 1989.

[16] M. E. Hernando, E. J. Gómez, F. del Pozo, and R. Corcoy. DIABNET: A qualitative model-based advisory system for therapy planning in gestational diabetes. *Medical Informatics*, 21:359–374, 1996.

[17] K. Kanazawa. *Reasoning about Time and Probability.* PhD thesis, Dept. of Computer Science, Brown University, 1992.

[18] U. Kjærulff. A computational scheme for reasoning in dynamic probabilistic networks. In *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence (UAI'92)*, pages 121–129, Stanford University, 1992. Morgan Kaufmann, San Francisco, CA.

[19] A. E. Nicholson and J. M. Brady. Sensor validation using dynamic belief networks. In *Proceedings of the 8th Conference on Uncertainty in Artificial Intelligence (UAI'92)*, pages 207–214, Stanford University, 1992. Morgan Kaufmann, San Francisco, CA.

[20] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Francisco, CA, 1988. Revised second printing, 1991.

[21] E. Santos and J. D. Young. Probabilistic temporal networks: A qualitative unified framework for reasoning with time and uncertainty. *International Journal of Approximate Reasoning*, 20:263–291, 1999.

[22] A. Y. Tawfik and E. Neufeld. Temporal reasoning and Bayesian networks. *Computational Intelligence*, 16(3):349–377, 2000.